HIGH SPECIFICITY MARKER DETECTION

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is related to U.S. Provisional patent application Serial Number 60/192,229 filed March 27, 2000, which is incorporated herein, in its entirety, by reference.

FIELD OF THE INVENTION

The present invention relates to methods of detecting tumor cells and evidence of cancer in samples.

BACKGROUND OF THE INVENTION

Rare circulating tumor cells in blood may be detected by amplifying mRNA of tumor- or tissue-specific markers utilizing RT-PCR. While RT-PCR can magnify target-specific nucleic acids up to 10^{20}-fold (Waldman *et al.*, 1998, Dis. Colon Rectum, 41:310-5; Cagir *et al.*, 1999, Ann. Intern. Med., 131:805-811; each of which is incorporated herein by reference), enhanced detection is associated with a high false positive rate (Burchill *et al.*, 1995, Br. J. Cancer, 71:278-281; Battaglia *et al.*, 1998, Bone Marrow Transpl., 22:693-698; Krismann *et al.*, 1995, J. Clin. Oncol., 13:2769-2775; each of which is incorporated herein by reference). This has been especially true in RT-PCR studies examining the ability to detect rare tumor cells in blood employing epithelial cell markers (Burchill *et al.*, 1995, *supra*; Battaglia *et al.*, 1998, *supra*; Krismann *et al.*, 1995, *supra*; Lopez-Guerrero *et al.*, 1997, Clin. Chim. Acta, 263:105-116; each of which is incorporated herein by reference). The high false positive rates appear to arise from illegitimate transcription of epithelial cell markers (Krismann *et al.*, 1995, *supra*; Lopez-Guerrero *et al.*, 1997, *supra*; Bostick *et al.*, 1998, J. Clin. Oncol., 16:2632-2640; Traweek *et al.*, 1993, Am. J. Pathol., 142:1111-1118; Hoon *et al.*, 1995, Cancer, 76:533-534;

Pelkey *et al.*, 1996, Clin. Chem., 42:1369-1381; Jung *et al.*, 1999, Br. J. Cancer, 81:870-873; each of which is incorporated herein by reference).

Illegitimate transcription (or ectopic transcription) is a general phenomenon of basal, very low level, transcription of any gene in any cell type, and has been well documented in many tissue-specific genes (Chelley *et al.*, 1991, J. Clin. Invest., 88:1161-1166; Chelley *et al.*, 1989, Proc. Natl. Acad. Sci. USA, 86:2671-2621; each of which is incorporated herein by reference). Importantly, detection techniques which first amplify transcription of spliced mRNA by such means as PCR are able to detect the expression of a tissue-specific gene from very minute amounts of mRNA present in a "non-expressing" cell sample. Therefore, illegitimate transcription is particularly problematic when using such techniques if the assay is sensitive enough to detect the illegitimate transcription product.

The extremely high sensitivity of RT-PCR has revealed that cells in the blood illegitimately transcribe genes that were previously considered markers of specific epithelia (Sarkar & Sommer, 1989, Science, 244:331-334; Chelley *et al.*, 1989, *supra*; Negrier *et al.*, 1998, Br. J. Haematol., 100:33-39; Gala *et al.*, 1998, Clin. Chem., 44:472-481; Champelovier *et al.*, 1999, Anti-Cancer Res., 19:2073-2078; each of which is incorporated herein by reference). Transcripts for PSA and PSM (prostate cancer), CK-19 and CK-20 (gastric, colon, and breast cancer), CEA (colorectal cancer), CK-18 (breast cancer), CK-8 (breast cancer), MUC-1 (breast, ovary, colon, and lung cancer), and GA733.2, (breast cancer) have been detected in peripheral blood from healthy volunteers (Burchill *et al.*, 1995, *supra*; Lopez-Guerrero *et al.*, 1997, *supra*; Bostick *et al.*, 1998, *supra*; Traweek *et al.*, 1993, *supra*; Hoon *et al.*, 1995, *supra*; Gala *et al.*, 1998, *supra*; Champelovier *et al.*, 1999, *supra*). Illegitimate transcription and background mRNA expression in normal blood limit the sensitivity of detecting circulating cancer cells employing PSM, CEA, and CK-18 (Zippelius *et al.*, 1997, J. Clin. Oncol., 15:2701-2708, which is incorporated herein by reference). Whether every cell has the ability to generate ectopic transcripts, and if ectopic transcripts have a biological role is unknown (Cooper *et al.*, 1994, Ann. Med., 26:9-14; Sommer & Sarkar, 1989, Science, 245:261; each of which is incorporated herein by reference). Since illegitimately transcribed mRNA levels are extremely low, it is unlikely that a biological role would involve protein synthesis.

Illegitimate transcription may represent the cost to cells of completely inactivating the transcription of thousands of "leaky" genes (Cooper *et al.*, 1994, *supra*).

Colorectal cancer is the third leading cause of cancer and cancer-related mortality worldwide (Pihl *et al.*, 1981, J. Surg. Oncol., 16:333-341; Toribara & Sleisenger, 1995, New Eng. J. Med., 332:861-7; Larson *et al.*, 1986, Arch. Surgery, 121:535-40; Silverberg *et al.*, 1990, Cancer Statistics, 40:9-26; Greenwald, 1992, Cancer 70(Suppl. 5):1206-1215; Cresanta, 1992, Prim. Care, 19:419-441; Jessup *et al.*, 1996, Cancer, 78:918-926; each of which is incorporated herein by reference). Forty percent of patients believed to be cured by surgery suffer disease recurrence within 3 years. At present, there are no effective blood-based methods to detect post-operative disease recurrence and reduce cancer-related mortality. Thus, clinical outcomes in patients with colorectal cancer could be substantially improved by the availability of more sensitive and specific diagnostic markers for post-operative surveillance (Shapiro, 1992, Cancer, 75(Suppl. 5):1252-1258; Smart, 1992, Cancer, 75(Suppl. 5):1246-1251; each of which is incorporated herein by reference).

Guanylyl cyclase C (GC-C) (also known as the ST or *E. coli* heat-stable enterotoxin receptor) is a cell surface receptor which mediates fluid and electrolyte secretion, with expression restricted to brush border membranes of intestinal mucosa cells from the duodenum to the rectum, and which is not expressed by extra-intestinal tissues (Gyles, 1971, Ann. N.Y. Acad. Sci., 176:314-322; Dejonge, 1975, FEBS Lett., 53:237-242; Guarino *et al.*, 1987, Dig. Dis. Sci., 32:1017-1026; Almenoff *et al.*, 1993, Mol. Microbiol., 8:865-873; Guerrant *et al.*, 1980, J. Infect. Dis., 142:220-228; Carrithers *et al.*, 1994, Gastroenterology, 107: 1653-1661; Krause *et al.*, 1994, Gut, 35:1250-1257; Cohen *et al.*, 1988, Gastroenterology, 94:367-373; Guarino *et al.*, 1987, Pediatric. Res., 21:551-555; each of which is incorporated herein by reference). GC-C expression persists after intestinal mucosal cells undergo neoplastic transformation and is expressed by primary and metastatic colorectal tumors regardless of their anatomical location (Huott *et al.*, 1988, J. Clin. Invest., 82:514-523; Guarino *et al.*, 1987, Am. J. Physiol., 253:G775-G780; Cohen *et al.*, 1993, J. Cell. Physiol., 156:138-144; Mann *et al.*, 1993, Am. J. Physiol., 264:G172-G178; each of which is incorporated herein by reference). GC-C is not expressed by tumors originating from outside the alimentary/gastrointestinal tract (Guerrant *et al.*, *supra*; Carrithers *et al.*, 1994, Gastroenterology, *supra*; Krause *et al.*, supra; Cohen *et al.*,

supra, Guarino *et al.*, 1987, Pediatric Res., *supra*.; Carrithers *et al.*, 1996, Proc. Natl. Acad. Sci. USA, 93:14827-14832; each of which is incorporated herein by reference). These data suggest that GC-C may be a unique marker for detecting metastatic colorectal cancer cells in blood during post-operative surveillance (Waldman *et al.*, 1998, Dis. Colon Rectum, *supra*; Cagir *et al.*, 1999, *supra*.; Carrithers *et al.*, 1996, Dis. Colon Rectum, 39:171-181, which is incorporated herein by reference).

Previous studies in colorectal cancer patients have demonstrated that GC-C can identify micrometastatic foci in lymph nodes evaluated as free of disease by standard histopathology. Importantly, detection of micrometastases by GC-C RT-PCR was associated with a greatly enhanced risk of colorectal cancer-related mortality. GC-C analysis may be a sensitive and specific method for detecting clinically significant colorectal cancer micrometastases in lymph nodes, and could improve the accuracy of staging.

Similarly, analyzing GC-C expression in blood to detect rare circulating colorectal tumor cells could improve the early detection of disease recurrence in patients undergoing post-operative surveillance. Current surveillance paradigms have not improved the overall survival of patients with recurrent colorectal cancer, in part, reflecting their inability to detect recurrence at a point amenable to intervention (Virgo *et al.*, 1995, JAMA, 273: 837-1841; Wade *et al.*, 1996, J. Am. Coll. Surg., 182:353-361; Moertel *et al.*, 1993, JAMA, 270:943-947; Schiessel *et al.*, 1986, Brit. J. Surg., 73:342-344; Bohm *et al.*, 1993, Dis. Colon Rectum, 36:280-286; Nelson, 1995, Sem. Oncol., 22:488-493; each of which is incorporated herein by reference). Development of a more effective surveillance marker would have significant impact on the management and outcome of colorectal cancer. Preliminary studies detected GC-C mRNA in blood from colorectal cancer patients, although there was no obvious correlation between the detection of this transcript and disease stage (Carrithers *et al.*, 1996, Proc. Natl. Acad. Sci. USA, *supra*; Bustin *et al.*, 1999, Br. J. Cancer, 79:1813-1820, which is incorporated herein by reference). In addition, GC-C mRNA was detected in the blood of some healthy volunteers (Bustin *et al.*, 1999, *supra*). Like other epithelial cell markers, GC-C may undergo illegitimate transcription in blood that may undermine its utility for post-operative surveillance.

Most paradigms for post-operative colon cancer surveillance include repeated

measurements of serum carcino-embryonic antigen (CEA) (Wamego *et al.*, 1978, Ann. Surg., 188:481-493; Sugarbaker *et al.*, 1976, Cancer, 38:2310-2315; Boey *et al.*, 1984, World J. Surg., 8:279-286; Northover, 1986, Gut, 27:117-121; each of which is incorporated herein by reference). Analysis of expression of GC-C by RT-PCR may be more sensitive and specific than CEA as a marker for metastatic colorectal cancer in blood. Whereas CEA is produced by <80% of colorectal tumors, GC-C has been detected in all primary and metastatic colorectal tumors examined. While CEA is expressed by some extra-intestinal tumors, GC-C is expressed only by colorectal tumors. CEA is expressed by tissues other than intestine that are involved in non-neoplastic conditions while GC-C has been identified only in colorectal cancer cells outside the intestine. In a retrospective analysis, GC-C was identified in lymph nodes of all patients who were node-negative by histopathology and who developed recurrent disease whereas CEA was identified in lymph nodes of only one of those patients (Cagir *et al.*, 1999, *supra*).

There is a need for improved methods of detecting the presence of metastasized cancers, including colon cancer. There is a need for methods of reducing the background signals caused by illegitimate transcription of cell markers used for the detection of cells that have migrated from their normal location in the body, including metastatic cancer cells. In particular there is a need to improve the accuracy and to decrease false-positive signals in highly sensitive, mRNA detection assays.

SUMMARY OF THE INVENTION

This invention provides methods of detecting the presence of a disseminated cell marker in a sample comprising the steps of eliminating illegitimate transcription-positive cells from the sample, and detecting the presence of mRNA that encodes the marker. The invention also provides methods of diagnosing metastatic cancer comprising detecting the presence of a disseminated cell marker for cancer cells identified as from the primary cancer in a sample that does not normally express said marker, comprising the steps of eliminating illegitimate transcription-positive cells from the sample, and detecting the presence of mRNA that encodes the marker. The invention also provides methods of detecting the presence of a tissue-specific marker in a sample not associated with the expression of the tissue-specific marker, comprising the steps of eliminating CD34+ cells

from the sample, and detecting the presence of mRNA encoding the tissue-specific marker. The invention also provides methods of detecting the presence of a disseminated cell in a sample, comprising the steps of eliminating CD34+ cells from the sample, and detecting the presence of mRNA that encodes a marker associated with the disseminated cell. The invention further provides kits for detecting the presence of a disseminated cell marker in a sample and kits for detecting the presence of a disseminated cell marker for metastatic cancer cells, comprising an affinity column; and primers for detecting the presence of mRNA encoding the marker.

These and other aspects of the invention are described more fully below.

All publications, patents and patent applications cited herein are hereby incorporated by reference in their entirety.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a photograph of an ethidium bromide-stained gel containing the products of nested RT-PCR analysis of GC-C expression in blood mononuclear cells from representative samples of healthy volunteers (n=20) and Dukes' stage D patients (n=24). Total RNA (1 μg) extracted from mononuclear cells of healthy volunteers and Dukes' stage D patients was subjected to nested RT-PCR employing GC-C-specific primers. T84 colorectal carcinoma cells served as a positive control for GC-C expression. H_2O (negative control) indicates no input RNA. Molecular weight markers are shown in the left-most lane of the gel; sizes are shown to the left of the gel. The arrow at the right indicates the size of the human GC-C RT-PCR product (~250 bp) predicted from the defined sequence.

Figures 2A and 2B present photographs of ethidium bromide-stained gels showing nested RT-PCR products, representing expression of GC-C transcripts in purified blood mononuclear cells. Total RNA (1 μg) extracted from (Figure 2A) plasma, granulocytes, erythrocytes, platelets, and mononuclear cells, or (Figure 2B) purified monocytes, T cells, NK cells, B cells, and CD34+ progenitor cells, and analyzed by nested RT-PCR employing GC-C specific primers. T84 cells served as a positive control for GC-C expression. H_2O (negative control) indicates no input RNA. Molecular weight markers and their sizes are indecated the left-most lanes of each gel. The arrows at the right of

each gel indicate the size of the human GC-C RT-PCR product (~250 bp) predicted from the defined sequence.

Figure 3 is a photograph of an ethidium bromide-stained gel showing nested RT-PCR products, representing the expression of GC-C mRNA in peripheral blood mononuclear cells enriched in CD34+ cells from a patient undergoing peripheral blood progenitor cell harvest following treatment with Neupogen. For PBPC mobilization in preparation for bone marrow ablation and autologous transplantation, a patient with breast cancer received 10 μg/Kg/day (total daily dose of 600 μg) of Neupogen™ (G-CSF; Amgen; Thousand Oaks, CA) as an IV bolus for 3 consecutive days. Leukaphoresis was performed on days 3 and 6 following the first dose of Neupogen™. Total RNA (1 μg) extracted from the mononuclear cells obtained on days 3 and 6 was subjected to RT-PCR employing GC-C-specific primers. The patient had a leukocyte count of 107,000/μl on day 3 and 17,000/μl on day 6. T84 cells served as a positive control for GC-C expression. H_2O (negative control) indicates no input RNA. Molecular weight markers and their sizes are indicated in the left lane. The arrow at the right indicates the size of the human GC-C RT-PCR product (~250 bp) predicted from the defined sequence.

Figure 4 is a photograph of an ethidium bromide-stained gel showing nested RT-PCR products, representing the expression of GC-C transcripts following depletion of CD34+ progenitor cells from peripheral blood mononuclear cells. Mononuclear cells were depleted of CD34+ progenitor cells as described in Example 1. Total RNA was extracted from mononuclear cells before and after depletion of CD34+ progenitor cells, and 1 μg was subjected to nested RT-PCR employing GC-C specific primers. T84 cells served as a positive control for GC-C expression. H_2O (negative control) indicates no input RNA. Molecular weight markers and their sizes are indicated in the right lane. The arrow indicates the predicted size of the human GC-C RT-PCR product (~250 bp).

Figures 5A and 5B present photographs of ethidium bromide-stained gels showing nested RT-PCR products, representing the expression of epithelial cell biomarkers in CD34+ progenitor cells. Total RNA (1 μg) extracted from 10^6 CD34+ progenitor cells was subjected to nested RT-PCR analysis employing epithelial cell marker-specific primers. In Figure 5A, the arrows indicate the defined sequence-predicted sizes of the RT-

PCR products for human PSA (~335 bp), PSM (~200 bp), GC-C (~250 bp), and CEA (~162 bp) and GC-C (~250 bp). In Figure 5B, the arrows indicated the defined sequence-predicted sizes of the RT-PCR products for human MUC-1 (~350 bp), CK-19 (~460 bp), CK-20 (~370 bp), GA733.2 (~700 bp) and CEA (~162 bp). Total RNA extracted from prostate was employed as a positive control for PSA and PSM RT-PCR analysis (gel to left of Figure 5A). T84 cells served as a positive control for expression of the other epithelial biomarkers. H_2O (negative control) indicates no input RNA. Molecular weight markers appear in the left-most lanes of the gels, and their sizes are indicated to the left of the gels.

Figure 6 presents photographs of two ethidium bromide-stained gels showing nested RT-PCR products, representing the expression of epithelial cell biomarkers in mononuclear cells depleted of CD34+ progenitor cells. Mononuclear cells were depleted of CD34+ progenitor cells by column chromatography as described in Example 1. Total RNA (1 μg) was extracted from pre-column (pre MNC) and post-column (post MNC) (depleted of CD34+ progenitor cells) mononuclear cells, and was subjected to nested RT-PCR, employing primers specific for β-actin, GC-C, CEA, CK-19, CK-20, and MUC-1. T84 cell RNA provided the positive controls. Molecular weight markers and their sizes are indicated in the left lanes.

Figure 7 presents photographs of two ethidium bromide-stained gels showing nested RT-PCR products, representing the expression of illegitimate transcripts of GC-C and CEA in the blood of healthy volunteers. Total RNA extracted from mononuclear cells of healthy subjects (n=20) was serially diluted and subjected to nested RT-PCR employing both GC-C- and CEA-specific primers. Quantities of input RNA are indicated along the top of the gels. The arrows indicate the sizes of the RT-PCR products of human GC-C (~250 bp) and CEA (~162 bp), predicted from their defined sequences. T84 cells served as a positive control for CEA and GC-C expression. H_2O (negative control) indicates no input RNA. Molecular weight markers and their sizes are indicated in the left lanes of the gels. These data are representative of samples examined from 20 healthy volunteers.

Figure 8A presents photographs of ethidium bromide-stained gels showing the threshold for detecting transcripts of GC-C and CEA in blood from Dukes' Stage D patients. Total RNA, extracted from mononuclear cells of Dukes' stage D patients, was

serially diluted and subjected to nested RT-PCR employing both GC-C- and CEA-specific primers. The RT-PCR products presented in the gels of Figure 8A are representative of the results with samples from the 24 Dukes' stage D patients examined. T84 cells served as a positive control for CEA and GC-C expression. H_2O (negative control) indicates no input RNA. Quantities of input RNA are indicated along the tops of the gels. The arrows indicate the sizes of the RT-PCR products for human GC-C (~250 bp) and CEA (~162 bp) predicted from their defined sequences. Molecular weight markers and their sizes are indicated in the left lanes of the gels. Figure 8B presents a graph comparing the sensitivity of RT-PCR employing GC-C- and CEA-specific primers to detect circulating tumor cells in Dukes' stage D patients. Total RNA was extracted from mononuclear cells of Dukes' stage D patients, and was serially diluted and subjected to nested RT-PCR employing both GC-C- and CEA-specific primers, as for Figure 8A. The percentage of samples (patients) which yielded GC-C- or CEA-specific amplicons was calculated for each quantity of total RNA analyzed.

Figure 9 presents photographs of two ethidium bromide-stained gels showing the sensitivity of nested RT-PCR employing GC-C-specific primers to detect human colorectal cancer cells in blood. Total RNA was extracted from the indicated numbers of mononuclear blood cells spiked with a single T84 cell (~200 copies of GC-C mRNA) or a single Caco2 cell (~20 copies of GC-C mRNA). Total RNA (0.5 μg) was subjected to nested RT-PCR analysis employing GC-C-specific primers. T84 and Caco2 cells served as respective positive controls for GC-C expression. H_2O (negative control) indicates no input RNA. Molecular weight markers and their sizes are indicated in the left lanes of the gels. The arrows indicate the size of the RT-PCR product for human GC-C (~250 bp) predicted from the defined sequence.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Nucleic acid-based diagnostic testing has been plagued by the presence of background levels of the disease marker of interest. Since these levels are low and often spurious, it has been referred to as "illegitimate transcription," with the inference that all cells have leaky transcription and small levels of all transcripts are produced constitutively. It has been discovered that this low level transcription may be ascribed to

distinct populations of stem cells, producing discrete levels of transcript per cell. One such cell type (a hematopoeitic stem cell expressing the surface glycoprotein CD34; CD34+ stem cells (Kronenwett *et al.*, 2000, Stem Cells, 18:320-330, which is incorporated herein by reference)) has been demonstrated to be the source of this background for a number of markers.

The present invention arises out of the discovery that CD34+ progenitor cells illegitimately transcribe a variety of epithelial cell-specific markers including GC-C, prostate-specific antigen (PSA), prostate-specific membrane antigen (PSM), carcinoembryonic antigen (CEA), cytokeratin-19 (CK-19), cytokeratin-20 (CK-20), mucin 1 (MUC-1), and gastrointestinal-associated antigen (GA733.2). CD34+ cells are the source of the high false positive rate generally observed when epithelial cell-specific markers are employed to detect rare circulating metastatic cancer cells by RT-PCR. Background signals reflect low-level transcription of these markers in CD34+ cells, and depletion of CD34+ cells or limiting the quantity of RNA analyzed can reliably eliminate false positive results.

The removal, destruction, or modification of CD34+ cells will improve assays for markers based on analysis of any human or veterinary tissue. The limitations to the utility of epithelial cell markers for detecting rare circulating tumor cells can be alleviated either by separating CD34+ and tumor cells (positive or negative purification) prior to RNA extraction or limiting the amount of total RNA analyzed to that below the limit of detection of illegitimate transcripts, prior to RT-PCR.

GC-C undergoes illegitimate transcription by blood mononuclear cells producing a high false positive rate in healthy volunteers. Illegitimate transcription of GC-C was localized specifically to CD34+ progenitor cells, which were the source of false positive signals for seven other epithelial cell markers. The illegitimately transcribed GC-C in CD34+ cells falls below the minimum threshold for expression of functional receptors on the cell surface (Waldman *et al.*, 1998, Cancer Epid. Bio. Prev., 7:505-514, which is incorporated herein by reference). Background signals reflect low-level transcription of these markers and depletion of CD34+ cells or limiting the quantity of RNA analyzed can reliably eliminate false positive results. Employing this technique, GC-C expression was detected in the circulation of all patients with metastatic colorectal cancer, but not in any

healthy volunteer examined. Thus, analysis of GC-C expression by RT-PCR is a sensitive and specific diagnostic tool for early detection of disease recurrence in patients who have undergone resection for colorectal cancer. Similarly, other epithelial cell markers can be useful for detecting rare circulating tumor cells following elimination of signals reflecting illegitimate transcription by separating CD34+ cells from tumor cells or limiting the quantity of RNA analyzed.

CEA expression was specifically detected in blood of <30% of patients with metastatic colorectal cancer compared to GC-C, which was detected in blood from all of those patients. In addition, GC-C in blood from patients with metastatic colorectal cancer was detected by RT-PCR employing quantities of total RNA as low as 0.08 μg while CEA was detected with no less than 0.5 μg of RNA. These data suggest that GC-C is more frequently expressed and more abundant than CEA in colorectal cancer cells. As a result, GC-C may be a more sensitive and specific biomarker than CEA for detection of rare metastatic colorectal cancer cells in blood.

The practice of the present invention employs, unless otherwise indicated, conventional methods of immunology, microbiology, molecular biology and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See, *e.g.*, Sambrook *et al.*, eds., Molecular Cloning: A Laboratory Manual (2nd ed.) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989); Ausubel *et al.*, eds., Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY (2000); Glover, ed., DNA Cloning: A Practical Approach, Vols. I & II; Colowick & Kaplan, eds., Methods in Enzymology, Academic Press; Weir & Blackwell, eds., Handbook of Experimental Immunology, Vols. I-IV, Blackwell Scientific Pubs. (1986); Coligan *et al.*, eds., Current Protocols in Immunology, John Wiley & Sons, New York, NY (2000), each of which is incorporated herein by reference.

Various definitions are made throughout this document. Most words have the meaning that would be attributed to those words by one skilled in the art. Words specifically defined either below or elsewhere in this document have the meaning provided in the context of the present invention as a whole and as typically understood by those skilled in the art.

As used herein, the term "disseminated" in reference to a cell means a cell that is

found in a location in the body that is different from its site of origin or normal location in the body. By way of non-limiting example, a malignant melanoma cell found in the peripheral blood of an individual would be a disseminated cell because its site of origin or normal location is the skin and it has been found in different site, in this case, the blood. A colon cancer cell found in the bone marrow of an individual is disseminated from its site of origin, the colon. Cancer cells that have metastasized are, by definition, disseminated cells because they have spread from their site or tissue of origin to a different site in the body. A disseminated cell can also be a cell that has begun to inappropriately express proteins, which may be indicative of a disease state.

As used herein, the term "disseminated cell marker" refers to a gene product associated with a particular cell or tissue type that may serve as an indication that a cell has become disseminated from its site of origin or normal location in the body. Any cell- or tissue-specific marker (also called differentiation specific antigens) can be a "disseminated cell marker" if that marker is found in a region or site of the body where that cell- or tissue-specific marker is not typically expected to be found. The presence of a disseminated cell marker is indicative of the presence of a disseminated cell. Depending upon the sample examined, many cell or tissue type markers can serve as disseminated cell markers. If the sample examined is the blood, then any cell or tissue type marker that is not expressed by cells that are normally to be found in the blood, can serve as a disseminated cell marker. Examples of markers that can be used as disseminated cell markers include, but are not limited to, guanylyl cyclase C (GC-C), Cdx-1, Cdx-2, sucrase isomaltase, lactase, carbonic anhydrase, prostate specific antigen (PSA), prostate specific membrane antigen (PSM), cytokeratin 18 (CK-18), cytokeratin 19 (CK-19), cytokeratin 20 (CK-20), carcinoembryonic antigen (CEA), ErbB2, Erb-B3, epithelial mucin-1 (MUC-1), epithelial mucin-18 (MUC-18), gastrointestinal tumor associated antigen 733.2 (GA 733.2), desmoplakin I (DPL I), epithelial glycoprotein 40 (EGP-40), tyrosinase, thyroglobulin (TGB), tyrosine hydroxylase, and neuron-specific glycoprotein (NPGP 9.5).

In one aspect of the invention, the disseminated cell is a cancer cell and the disseminated cell marker is a differentiation specific antigen associated with the tissue of origin of the tumor cell. The presence of such differentiation specific antigens in samples of tissue or body fluids, that are distinct from the tissue of origin of the tumor, is indicative

of the presence of disseminated cells of the tumor in that tissue or body fluid. Methods related to the detection of rare cancer cells, in particular cancer cells that have become disseminated from their site of origin, are described in Doeberitz & Lacroix, 1999, Cancer Metastasis Rev., 18:43-64, which is incorporated by reference in its entirety; the references cited therein are incorporated by reference in their entirety.

The eliminating step can be accomplished by the removal, destruction, or modification of the cells that are positive for illegitimate transcription.

As used herein, the term "illegitimate transcription-positive" in reference to a cell, refers to a cell that inappropriately expresses a gene, generating some low level amount of mRNA transcript encoding a particular protein product. A cell inappropriately expresses a gene when that cell is not of the type of cell normally associated with the expression of that gene. For example, cells in the blood do not normally express the gene for tyrosinase, which is an enzyme specifically expressed in melanocytes or in tumor cells derived from melanocytes (meloanoma). If a blood cell was expressing the tyrosinase gene and generating mRNA transcript encoding tryrosinase, it would be an illegitimate transcription-positive cell. Cells that are illegitimate transcription-positive for a particular marker are cells that do not normally express the marker protein, *i.e.*, cells which are not identified with such marker expression. It has been discovered that CD34+ cells are illegitimate transcription-positive for many differentiation specific markers associated with other cell types. In preferred embodiments of the invention, illegitimate transcription-positive cells are CD34+ cells.

As used herein, the term "sample" includes any material, such as bodily fluids or portions of tissue, collected from an individual. The sample can be any tissue or body fluid. By way of non-limiting example, samples may include any of the following tissues or fluids: blood, bone marrow, sputum, semen, stool, gastric fluid, gastric juices, alimentary canal juices, saliva, urethral secretions, vaginal secretions, lung, peritoneal or pericardial lavage, urine, lymph, and cerebro-spinal fluid (CSF).

Samples comprise illegitimate transcription-positive cells, particularly CD34+ cells. Examples of such preferred samples include blood and bone marrow. Those skilled in the art will appreciate that the methods of the invention are particularly applicable to samples that comprise illegitimate transcription-positive cells, particularly cells that are

CD34+.

As used herein, the term "detecting" in reference to mRNA that encodes a marker includes any method of analysis that demonstrates the presence of mRNA encoding the marker of interest. The detecting step can be accomplished by any method that identifies the presence of a mRNA transcript. Thus, detection of a mRNA transcript that encodes a disseminated cell marker can be accomplished by, for example, PCR, RT-PCR, or antibody-based methods following the translation of the mRNA transcript. Detecting includes direct detection of the presence of a particular mRNA, and indirect detection of a particular mRNA by detection of a cDNA product or a protein product of that mRNA. Detecting includes the use of PCR-based methods for demonstrating the presence of mRNA transcripts, including but not limited to, direct PCR analysis of mRNA extracts and RT-PCR, using marker-specific oligonucleotide primers. Detecting also includes immunologically-based techniques to demonstrate the presence of the protein product of the mRNA of interest, following *in vitro* translation.

As used herein, a "CD34+ cell" can be any cell that expresses a part of the CD34 glycoprotein. CD34 is a 115 kD glycosylated Type I transmembrane protein, mainly expressed in precursors of hematopoietic cells and in the vascular endothelium. CD34 expression has been found in 1 – 4 % of adult bone marrow mononuclear cells (including marrow-repopulating cells, all multipotent and committed myeloid progenitors, B and T lymphoid precursors, osteoclast precursors, and most likely the precursors for stromal cells), and in less than 1% of peripheral blood mononuclear cells. In non-lymphohemopoietic tissues its expression is has been identified in endothelial cells and in some cells of the skin (Silvestri *et al.*, 1992, Haematologica, 77:265-273, which is herein incorporated by reference).

CD34+ cells can be removed, modified or destroyed by a variety of methods, which include affinity technologies, physical separation technologies or chemical technologies. Some examples of physical separation technologies would be ultrasound based acoustic levitation, field flow fractionation and separations based on charge, rigidity, aggregation, density or sensitivity to electromagnetic radiation. Affinity technologies can be based on molecules such as antibodies, partial antibodies, antibody fragments, modified antibodies, bacteria or viruses displaying peptides or proteins with

affinity to CD34+ cells, carbohydrates, peptides, nucleic acids or lipids with affinity for CD34+ cells. Chemical technologies would involve methods that either destroy or modify the ability of CD34+ cells to produce or contain background levels of markers used in diagnostic tests. These technologies can be comprised of treatment of the cells with chemical or biological substances or treatment of the cells with physical, chemical or biological methods such that the chemistry of the cell is modified. This modification would serve to remove the background level of diagnostic markers either by adjusting their levels or by removing them.

In preferred embodiments, CD34+ cells are removed from a blood sample using an antibody based affinity process.

As used in this specification and the appended claims, the singular forms "a," "an" and "the" include plural references unless the content clearly dictates otherwise. Thus, for example, reference to "a cell" includes a mixture of two or more cells.

The present invention relates to method of detecting a disseminated cell marker in a sample. In one aspect of the present invention, a disseminated cell marker is indicative of a metastatic cancer cell, *i.e.*, a cell from a tumor that has traveled to a site in the body that is distinct from its site of origin. Such a cell can be detected in a site that is distinct from its site of origin because of its expression of genes reflecting its particular tissue type or state of differentiation. The invention provides methods of detecting cancer cells, in particular cancer cells that have spread from their site of origin to other sites of the body, by examination of samples taken from sites in the body that are distinct from the original site of the tumor.

A variety of cancers that have disseminated from their site of origin can be detected by identifying transcripts for differentiation specific antigens associated with the tissue type of origin or oncofetal antigens inappropriately expressed by the cancer cell. A disseminated cancer cell can be identified by detecting the presence of mRNA transcripts for any protein that is associated with the cancer cell but that would not be expected to be expressed in the site of the sample being tested. Differentiation specific antigens and oncofetal antigens represent possible disseminated cell markers for identifying disseminated cancer cells. Such markers include, but are not limited to, PSA and PSM (prostate cancer), CK-19 and CK-20 (gastric, colon, and breast cancer), CEA (colorectal

cancer), CK-18 (breast cancer), CK-8 (breast cancer), MUC-1 (breast, ovary, colon, lung, and thyroid cancer), and GA733.2, (breast cancer), thyroglobulin (thyroid cancer), tyrosinase (melanoma (skin cancer)), β-HCG (testicular cancer), alpha-feto protein (AFP) (hepatocellular carcinoma (liver cancer)), Cdx1 and Cdx2 (colon, esophageal, and stomach cancer), and sucrase-isomaltase (colon, esophageal, and stomach cancer).

Those of skill in the art will recognize that the methods of the present invention are applicable to the detection of that detection of any disseminated cancer cells where a marker can be assigned to the cancer cell that would not be expected to be seen in a tested sample. The samples that are examined for evidence of disseminated cells can be from any tissue or body fluid. As the methods are directed to finding disseminated cells, appropriate sources for a sample will be prescribed by what is known about the site of origin of the disseminated cell of interest. By way of non-limiting example the sample may be from such tissues and fluids as blood, bone marrow, sputum, semen, stool, gastric, vaginal, lung, peritoneal or pericardial lavage, urine, lymph, cerebro-spinal fluid (CSF).

The present invention relates to methods of detecting a disseminated cell marker in a sample by eliminating cells which illegitimately transcribe genes. The elimination of illegitimate transcription-positive cells will improve the accuracy of assays designed to identify the presence of mRNA for a disseminated marker, by eliminating the cells which are not actually disseminated, but give false-positive signals of expression of the marker. Elimination of illegitimate transcription-positive cells will improve the reliability of these assays, so that detection of the presence of mRNA encoding a disseminated cell marker in a sample is truly reflective of the presence of a disseminated cell.

In one embodiment of the invention, illegitimate transcription-positive cells are eliminated by the removal of CD34+ cells from the sample. A preferred means of removal of CD34+ cells is by use of affinity column chromatography, for example, the CD34 Progenitor Cell Isolation Kit™ (Miltenyi Biotec; Bergisch Gladbach, Germany) can be used to selectively remove CD34+ cells from a sample.

Another aspect of the present invention includes various methods of determining whether a sample contains disseminated cells by determining whether the sample contains mRNA that encodes a disseminated cell marker. Detection of the presence of the mRNA is carried out by use of nucleotide sequence-based molecular analysis. Several different

methods are available for doing so including those using Polymerase Chain Reaction (PCR) technology, branched chain oligonucleotide hybridization technology, Northern blot technology, oligonucleotide hybridization technology, and *in situ* hybridization technology. The invention relates to oligonucleotide probes and primers used in the methods of identifying mRNA that encodes a disseminated cell marker. The mRNA sequence-based methods for determining whether a sample contains mRNA encoding a disseminated cell marker include but are not limited to polymerase chain reaction technology, branched chain oligonucleotide hybridization technology, Northern and Southern blot technology, *in situ* hybridization technology and oligonucleotide hybridization technology.

The methods described herein are meant to exemplify how the present invention may be practiced and are not meant to limit the scope of invention. It is contemplated that other sequence-based methodology for detecting the presence of specific mRNA that encodes a disseminated cell marker in samples may be employed according to the invention.

A preferred method for detecting mRNA that encodes a disseminated cell marker in genetic material derived from samples that do not normally express the disseminated cell marker uses PCR technology. PCR assays are useful for detecting mRNA encoding a disseminated cell marker in homogenized tissue samples and in body fluid samples.

PCR technology is practiced routinely by those having ordinary skill in the art and its uses in diagnostics are well known and accepted. Methods for practicing PCR technology are disclosed in "PCR Protocols: A Guide to Methods and Applications", Innis, M. A. *et al.*, eds., Academic Press, Inc., San Diego, CA (1990), which is incorporated herein by reference. Applications of PCR technology are disclosed in "Polymerase Chain Reaction" Erlich, H. A. *et al.*, eds. Cold Spring Harbor Press, Cold Spring Harbor, NY (1989), which is incorporated herein by reference. U.S. Pat. No. 4,683,202, U.S. Pat. No. 4,683,195, U.S. Pat. No. 4,965,188, and U.S. Pat. No. 5,075,216, which are each incorporated herein by reference, describe methods of performing PCR. PCR may be routinely practiced using, for example, the GeneAmp® Gold RNA PCR Reagent Kit from PE Biosystems (Foster City, CA).

PCR technology including RT-PCR allows for the rapid generation of multiple

copies of DNA sequences by providing sets of primers that hybridize to sequences present in an RNA or DNA molecule, and further by providing free nucleotides and an enzyme that fills in the complementary bases to the nucleotide sequence adjacent to and thereby between the primers with the free nucleotides to produce complementary strands of DNA. The enzyme will fill in the complementary sequences adjacent to the primers. If both of the primers hybridize to nucleotide sequences on the same small fragment of nucleic acid, exponential amplification of a specific double-stranded size product results. If only a single primer hybridizes to the nucleic acid fragment, linear amplification produces single-stranded products of variable length.

PCR primers can be designed routinely by those having ordinary skill in the art using sequence information. Many nucleotide sequences encoding a wide variety of tissue-specific markers which may serve as disseminated cell markers have been identified and can be found in the scientific literature and in such databases as GenBank.

The following are non-limiting examples of disseminated cell markers and their cognate nucleotide sequence entries in GenBank, which are incorporated herein by reference:

Marker	GenBank Accession No.
GC-C	U20230; P25092; AAC50381
Cdx1	U51095; NM001804; U15212; P47902; AAC50237; AAB40602
Cdx2	U51096; NM001265; Y13709; CAA74038; AAB40603
sucrase-isomaltase	NM001041; AAA60551; M22616; NP004659
PSA	XM008995
PSM	XM012114
CEA	NM001712; XM012777
MUC-1	AJ006206; AJ006205
GA733.2	NM002354
β-HCG	J00117; M38559; M54963
AFP	XM003498

The nucleotide sequences for a given disseminated cell marker of interest may be used to design primers that specifically amplify mRNA that encodes that disseminated cell marker. To perform this method, RNA is recovered from a sample, by standard extraction

techniques, and tested or used to make cDNA using well known methods and readily available starting materials. The RNA may be contained within the cells of the sample or may be in the non-cellular portion. The mRNA encoding a disseminated cell marker may be found in the extracellular portion of a sample, where, for example, necrosis results in the lysis of disseminated cells and the subsequent release of cell contents, including mRNA. Detection of the released mRNA would be indicative of the presence of disseminated cells in the sample.

Those having ordinary skill in the art can readily prepare PCR primers. A set of primers generally contains two primers. When performing PCR on extracted mRNA or cDNA generated therefrom, if the mRNA or cDNA encoding the disseminated cell marker is present, multiple copies of the mRNA or cDNA will be made. If it is not present, PCR will not generate a discrete detectable product. Primers are generally 8-50 nucleotides, preferably about 15-35 nucleotides, more preferably 18-28 nucleotides, that are identical or complementary to and therefor hybridize to the mRNA or cDNA generated therefrom which encodes a disseminated cell marker. In preferred embodiments, the primers are each 15-35 nucleotide fragments, more preferably 18-28 nucleotide fragments of the nucleic acid molecule that comprises the nucleotide sequence encoding a disseminated cell marker. The primer must hybridize to the sequence to be amplified. Typical primers are 18-28 nucleotides in length and generally have 50% to 60% G+C composition. The entire primer is preferably complementary to the sequence it must hybridize to. Preferably, primers generate PCR products 100 base pairs to 2000 base pairs. However, it is possible to generate products of 5 kb to 10 kb and more. If mRNA is used as a template, the primers must hybridize to mRNA sequences. If cDNA is used as a template, the primers must hybridize to cDNA sequences.

The mRNA or cDNA is combined with the primers, free nucleotides and enzyme following standard PCR protocols. The mixture undergoes a series of temperature changes. If the mRNA or cDNA encoding the disseminated cell marker of interest is present, that is, if both primers hybridize to sequences, the molecule comprising the primers and the intervening complementary sequences will be exponentially amplified. The amplified DNA can be easily detected by a variety of well known means. If no mRNA or cDNA that encodes the disseminated cell marker of interest is present, no PCR

product will be exponentially amplified. The PCR technology therefore provides an extremely easy, straightforward and reliable method of detecting mRNA encoding a disseminated cell marker in a sample.

PCR products may be detected by several well known means. The preferred method for detecting the presence of amplified DNA is to separate the PCR reaction material by gel electrophoresis and stain the gel with ethidium bromide in order to visual the amplified DNA if present. A size standard of the expected size of the amplified DNA is preferably run on the gel as a control.

In some instances, such as when unusually small amounts of RNA are recovered and only small amounts of cDNA are generated therefrom, it is desirable or necessary to perform a PCR reaction on the first PCR reaction product. That is, if difficult to detect quantities of amplified DNA are produced by the first reaction, a second PCR can be performed to make multiple copies of DNA sequences of the first amplified DNA. A nested set of primers are used in the second PCR reaction. The nested set of primers hybridizes between sequences hybridized by the first set of primers.

Another method of determining whether a sample contains cells expressing a disseminated cell marker is by branched chain oligonucleotide hybridization analysis of mRNA extracted from a sample. Branched chain oligonucleotide hybridization may be performed as described in U.S. Pat. No. 5,597,909, U.S. Pat. No. 5,437,977, and U.S. Pat. No. 5,430,138, which are each incorporated herein by reference. Reagents may be designed following the teachings of those patents and sequence information for a given disseminated cell marker.

Another method of determining whether a sample contains cells expressing mRNA encoding a disseminated cell marker is by Northern Blot analysis of mRNA extracted from a sample. The techniques for performing Northern blot analyses are well known by those having ordinary skill in the art and are described in Sambrook, J. *et al.*, (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. mRNA extraction, electrophoretic separation of the mRNA, blotting, probe preparation and hybridization are all well known techniques that can be routinely performed using readily available starting material.

The mRNA is extracted using poly dT columns and the material is separated by

electrophoresis and, for example, transferred to nitrocellulose paper. Labeled probes made from an isolated specific fragment or fragments can be used to visualize the presence of a complementary fragment fixed to the paper. Probes useful to identify mRNA in a Northern Blot have a nucleotide sequence that is complementary to mRNA transcribed from the gene that encodes a given disseminated cell marker. Those having ordinary skill in the art could design such probes or isolate and clone a given disseminated cell marker gene or cDNA which can be used as a probe.

Northern blot analysis is useful for detecting mRNA encoding a disseminated cell marker in homogenized tissue samples and cells in body fluid samples. It is contemplated that Northern Blot analysis of mRNA extracted from a tissue or body fluid sample could be used to detect mRNA encoding a disseminated cell marker.

Another method of detecting the presence of mRNA encoding disseminated cell marker is by oligonucleotide hybridization technology. Oligonucleotide hybridization technology is well known to those having ordinary skill in the art. Briefly, detectable probes which contain a specific nucleotide sequence that will hybridize to nucleotide sequence of mRNA encoding a given disseminated cell marker. RNA or cDNA made from RNA from a sample is fixed, usually to filter paper or the like. The probes are added and maintained under conditions that permit hybridization only if the probes fully complement the fixed genetic material. The conditions are sufficiently stringent to wash off probes in which only a portion of the probe hybridizes to the fixed material. Detection of the probe on the washed filter indicate complementary sequences.

One having ordinary skill in the art can design probes that are fully complementary to disseminated cell marker mRNA sequences but not to genomic DNA. Hybridization conditions can be routinely optimized to minimize background signal by non-fully complementary hybridization.

Oligonucleotide hybridization techniques are useful for detecting mRNA encoding a disseminated cell marker in homogenized tissue samples and cells in body fluid samples. It is contemplated that oligonucleotide hybridization analysis of mRNA extracted from a tissue or body fluid sample could be used to detect mRNA encoding a disseminated cell marker.

The presence of mRNA that encodes a disseminated cell marker or cDNA

generated therefrom can be determined using techniques such as *in situ* hybridization. *In situ* hybridization technology is well known by those having ordinary skill in the art. Briefly, cells are fixed and detectable probes which contain a specific nucleotide sequence are added to the fixed cells. If the cells contain complementary nucleotide sequences, the probes, which can be detected, will hybridize to them.

One having ordinary skill in the art can design probes useful in *in situ* hybridization technology to identify cells that express mRNA that encodes a disseminated cell marker. The probes should be designed to be fully complementary to mRNA sequences but not to genomic sequences for the marker gene of interest. Hybridization conditions can be routinely optimized to minimize background signal by non-fully complementary hybridization. The probes are fully complementary and do not hybridize well to partially complementary sequences. The probes may be detected by fluorescence. A common procedure is to label the probe with a biotin-modified nucleotide, and then detect the probe with fluorescently tagged avidin. Cells are fixed and the probes are added to the genetic material. Probes will hybridize to the complementary nucleic acid sequences present in the sample. Using a fluorescent microscope, the probes can be visualized by their fluorescent markers. The probes may also be labeled for direct detection by incorporating radiolabeled nucleotides or nucleotides having detectable non-radioactive labels. Such probe detection systems are well known to those of skill in the art.

The presence of mRNA encoding a disseminated marker in a sample can be indirectly assayed by translation of the mRNA present in the sample, followed by detection of disseminated marker protein product using immunological assays with protein-specific antibodies. Such immunological assays include, but are not limited to, immunoprecipitation, immunoblotting, and immunohistochemistry. The antibodies can be visualized through a variety of detection techniques well known to the art. The antibodies are detectably labeled or detected using a labeled second antibody or protein A.

Other embodiments of the invention will be readily understood by those of skill in the art.

The invention is further illustrated by way of the following examples, which are intended to elaborate several embodiments of the invention. These examples are not

intended to, nor are they to be construed to, limit the scope of the invention. It will be clear that the invention may be practiced otherwise than as particularly described herein. Numerous modifications and variations of the present invention are possible in view of the teachings herein and, therefore, are within the scope of the invention.

EXAMPLES

Example 1: Materials and Methods.

Clinical specimens

Blood and tissue specimens were obtained from the hematology/oncology clinic under an Institutional Review Board-approved protocol (Control #98.0614) at Thomas Jefferson University Hospital (Philadelphia, Pennsylvania) and the Cooperative Human Tissue Network (Philadelphia). Healthy volunteers and Dukes' Stage D patients were informed about the study and asked to participate. After informed consent was obtained, each participant received a unique identification number that was recorded on blood samples and any acquisition forms. Blood (~16 cc) collected into Vacutainer®CPT™ tubes containing sodium heparin was centrifuged at 25°C for 15 minutes at 1700 rpm and the resulting mononuclear cell, red blood cell, and granulocytes fractions recovered for RNA extraction. In some experiments, whole blood was centrifuged at 1300 rpm at 4°C for 10 minutes, the resulting supernatant containing the platelet-rich plasma was centrifuged at 3000 rpm at 4°C for 10 minutes, and the platelet pellet was recovered for RNA extraction.

Isolation and purification of platelets

In order to isolate and purify platelets, whole blood was spun at 1300rpm at 4°C for 10 minutes. The supernatant was transferred into a new 15 ml conical tube. This platelet-rich plasma was centrifuged at 3000 rpm at 4°C for 10 minutes and the supernatant was discarded. The platelet pellet was resuspended in Tris Buffer pH 7.6/protease inhibitor solution. Pellet preps were then freeze thawed in liquid nitrogen and placed in a water bath (37°C) for 10 minutes. Platelets were homogenized and spun in an ultracentrifuge at 30,000 rpm at 4°C for 1 hour. The supernatant was discarded and the

pellet was resuspended in Tris buffer pH 7.6 and aliquoted into Eppendorf tubes until analysis.

Peripheral blood progenitor cell (PBPC) mobilization

G-CSF increases the quantity of CD34+ stem cells in the peripheral circulation. To examine the relationship between the quantity of circulating CD34+ stem cells and the level of illegitimate transcription of epithelial cell markers, blood was obtained from a patient with breast cancer undergoing PBPC mobilization in preparation for autologous transplantation. The patient received 10 μg/Kg/day (total daily dose of 600 μg) of G-CSF (Neupogen™; Amgen; Thousand Oaks, CA) as an IV bolus for 3 consecutive days. Leukaphoresis was performed on days 3 and 6 following the first dose of G-CSF.

Cell culture

T84 and Caco2 human colon carcinoma cells, obtained from the American Type Culture Collection (Manassas, VA), were grown to confluence and used as positive controls for GC-C mRNA in RT-PCR analyses (10). T84 and Caco2 cells were grown in media containing DMEM/F12 with 10% FBS and 1% Pen/Strep. Adherent cell lines were routinely passaged by trypsinization every 3 to 4 days.

Nucleic acid extraction

Total RNA was extracted with a modified version of the acid guanidinium thiocyanate/phenol/chloroform method employing TRI-REAGENT™ (MRC; Cincinnati, Ohio). The concentration, purity, and amount of total RNA were determined by ultraviolet spectrophotometery. Only samples exhibiting intact 28S and 18S ribosomal RNA were subjected to RT-PCR. All RNA preparations were stored in RNase-free water (Promega; Madison, WI) at –70°C until analysis.

RT-PCR

The expression of epithelial cell markers in blood cells was examined by RT-PCR employing transcript-specific primer sets (Table 1). Reverse transcription of total RNA (≤1 μg) was performed with 0.25 units/ul of AMV reverse transcriptase (Panvera; Madison, WI) and buffer containing 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 4 mM $MgCl_2$, 1 mM each of dATP, dCTP, dGTP, and dTTP, 1 unit/μl RNase inhibitor (Panvera; Madison, WI), and 1 μM of the appropriate antisense primer in a total volume of 20 μl. Thermal cycling proceeded for 1 cycle at 50°C for 30 minutes, 99°C for 5 minutes (to

inactivate reverse transcriptase), and 4°C for 5 minutes. The resultant cDNA was subjected to PCR in the same reaction tube and included 2.5 units of TaKaRa Taq polymerase (Panvera; Madison, WI) in 10 μl of: 10 mM Tris-HCl, 50 mM KCl, 2.5 mM $MgCl_2$, and 0.2 μM of the appropriate sense primer. Incubation and thermal cycling conditions were: 95°C for 2 minutes, 1 cycle; 94°C for 30 seconds, 58°C for 30 seconds, 72°C for 90 seconds, 35 cycles; 72°C for 5 minutes, 1 cycle. Following RT-PCR, samples were stored at -4°C? until analysis. Nested PCR (70 cycles) was performed employing 5% of the PCR product (DNA) and 2.5 units of TaKaRa Taq polymerase (Panvera; Madison, WI) in 100 μl of: 10 mM Tris-HCl, 50 mM KCl, 2.5 mM $MgCl_2$, and 0.2 μM of the appropriate sense primer. Incubation and thermal cycling conditions were: 95°C for 2 minutes, 1 cycle; 94°C for 30 seconds, 58°C for 30 seconds, 72°C for 90 seconds, 35 cycles; 72°C for 5 minutes, 1 cycle. Amplicons were separated by 4% Nusieve 3:1 agarose (FMC Bioproducts; Rockland, Maine) and visualized by ethidium bromide. Amplicon identity was confirmed at least once by DNA sequencing. RT-PCR was performed utilizing primers for β-actin on all samples to confirm the integrity of RNA. RNA extracted from T84 human colon carcinoma cells was employed as a positive control for GC-C mRNA. Negative controls included RT-PCR incubations that omitted RNA template. Primers employed for GC-C amplification span predicted intron-exon junctions, reducing the probability that amplification products reflect contaminating DNA templates.

Table 1. Primer Sequences for RT-PCR Amplification

Transcript (Reference)	Primers 5'→3'	SEQ ID NO:	Amplicon Size (bp)
β-actin	TGC-CATCCTAAAAGC-CAC[a]	1	220
(1)	GGAGACCAAAAGC-CTTCATAC[b]	2	
GC-C	GTTTCCTATTTCTCCCACGAACTC[a]	3	533
(1 - 3)	TTTCTTGGTGTCCACAGAGGTA[b]	4	
GC-C Nested	GGACCACAACAGGAAAAGCAATG[a]	5	262
(2, 3)	AGGCAAGACGAAAGTCTCGTTT[b]	6	
CEA	TCTGGAACTTCTCCTGGTCTCTCAGCTGG[a]	7	220
(4)	TGTAGCTGTTGCAAATGCTTTAAGGAAGAAGC[b]	8	
CEA Nested (4)	GGGCCACTGTCGGCATCATGAT	9	160
CK-19	AGGTGGATTCCGCTCCGGGCA[a]	10	460
(4)	ATCTTCCTGTCCCTCGAGCA[b]	11	
CK-20	CAGACACACGGTGAACTATGG[a]	12	370
(4, 5)	GATCAGCTTCCACTGTTAGACG[b]	13	
MUC-1	CGTCGTGGACATTGATGGTACC[a]	14	288
(4)	GGTACCTCTCACCTCCTCCAA[b]	15	
PSA	TACCCACTGCATCAGGAACA[a]	16	455
(6)	CCTTGAAGCACACCATTACA[b]	17	
PSA Nested (6)	ACACAGGCCAGGTATTTCAG[a]	18	335
	GTCCAGCGTCCAGCACACAG[b]	19	
PSM	GAATGCCAGAGGGCGATCTA[a]	20	441
(6)	TTCTAGGAGCTTCTGTGCATCATAGTATCC[b]	21	
PSM Nested (6)	AGGGGCCAAAGGAGTCATTCTCTACTCCGA[a]	22	186
	CTCTGCAATTCCACGCCTAT[b]	23	
GA733.2	CTACAAGCTGGCCGTAAACT[a]	24	700
(4)	GTCCTTGTCTGTTCTTCTGA[b]	25	

[1] Carrithers *et al.*, 1996, Proc. Natl. Acad. Sci. USA, 93:14827-32.

[2] Waldman *et al.*, 1998, Dis. Colon Rectum, 41:310-5.

[3] Cagir *et al.*, 1999, Ann. Intern. Med., 131:805-811.

[4] Bostick *et al.*, 1998, J. Clin. Oncol., 16:2632-2640.

[5] Champelovier *et al.*, 1999, Anti-Cancer Res., 19:2073-2078.

[6] Grasso *et al.*, 1998, Cancer Res., 58:1456-1459.

[a]Forward primer.

[b]Reverse primer.

Isolation of purified cell components from blood mononuclear cells

Monocytes, NK cells, T cells, CD19+ B cells, and CD34+ progenitor cells were obtained commercially (Bio-Whittaker; Charlotte, NC). Purified NK and T cells were generously provided by Dr. Bice Perussia, Kimmel Cancer Institute, Thomas Jefferson University. CD34+ progenitor cells were isolated from peripheral blood with the CD34 Progenitor Cell Isolation Kit™ (Miltenyi Biotec; Bergisch Gladbach, Germany). Similarly, populations of mononuclear cells were depleted of CD34+ cells by use of this kit. CD34+ progenitor cells were indirectly magnetically labeled using hapten-conjugated primary monoclonal antibody directed to CD34 and an anti-hapten antibody, coupled to MACS™ microbeads (Bio-Whittaker, Charlotte, NC). Magnetically labeled cells were purified and recovered in the magnetic field of a MACS™ separator (Bio-Whittaker, Charlotte, NC).

Miscellaneous

All reagents were of analytical reagent grade. Results are representative of at least three experiments. Values representing the mean ± SD were calculated using Microsoft Excel™.

Example 2: Subject Characteristics.

Volunteer ages ranged from 20 to 51 years of age (y) (32.9 ± 2.4 y), and patient ages ranged from 33 to 79 y (59.4 ± 2.7 y). There is an inverse relationship between age and the quantity of circulating CD34+ stem cells (36), suggesting that those cells contributed less to results obtained with patients compared to volunteers. There were no significant differences between the ages of female (range = 23-51 y; 30.7 ± 3.3 y) and male (range = 20-48 y; 35.1 ± 3.8 y) volunteers, or female (range = 33-79 y; 57.8 ± 2.7 y) and male (range = 40-78 y; 61.1 ±3.0 y) patients. Four female and one male patient were African American; all other patients were Caucasian. One female and three male volunteers were African American; all other volunteers were Caucasian. Disease characteristics of patients are outlined in Table 2. Twenty one of 24 patients had hepatic metastases, no patient had pulmonary metastases, and 5 patients had bone metastases. All patients were receiving chemotherapy that included 5'-fluorouracil and leucovorin during

this study.

Table 2. Patient Characteristics.

Patient	Age/Race/Sex[a]	Metastases[b]	Overall Survival[c]	Chemotherapy[d]
1	79BF	H	36	FU/Lv/CPT-11
2	55WM	H/Bn	34	FU/Lv/CPT-11/Ox
3	59WM	H	36	FU/Lv/CPT-11/Ox
4	69WM	H	60	FU/Lv/CPT-11/Ox
5	58WM	H/Bn	12	FU/Lv/CPT-11/Ox
6	61BF	H	36	FU/Lv/CPT-11/Ox
7	66WM	H/Spleen	24	FU/Lv/CPT-11/Ox
8	78WM	Bn	60	FU/Lv/CPT-11
9	70WF	H/Lung	72	FU/Lv/CPT-11
10	55WM	H	36	FU/Lv
11	41WF	Lung	24	FU/Lv/CPT-11
12	64WM	H	60	FU/Lv/CPT-11
13	49WF	H	24	FU/Lv/CPT-11/Ox
14	52WM	H	36	FU/Lv/CPT-11/Ox
15	71WM	H/Lung	60	FU/Lv/CPT-11/Ox
16	33WF	H	36	FU/Lv
17	78WF	H	92	FU/Lv/CPT-11/Ox
18	55WF	H/Lung	48	FU/Lv/CPT-11/Ox
19	71WF	H/Lung	48	FU/Lv/CPT-11/Ox
20	66BF	H/Lung	36	FU/Lv/CPT-11
21	66WM	H/Bn/Lung	24	FU/Lv/CPT-11/Ox
22	50BF	H/Lung	48	FU/Lv/CPT-11/Ox
23	41WF	H/Lung/Bn	60	FU/Lv/CPT-11/Ox
24	40WM	Lung	24	FU/Lv/CPT-11/Ox

[a]B, black; W, white; M, male; F, female.

[b]H, hepatic; P, pulmonary, Br, brain; Bn, bone.

[c]Months.

[d]FU, 5'-fluorouracil; Lv, leucovorin; CPT-11, irinotecan ; Ox, oxaliplatin.

Example 3: Nested RT-PCR Detects GC-C Expression in Mononuclear Cells from Volunteers and Patients.

Total RNA (1 μg) extracted from mononuclear cells of 20 healthy volunteers and 24 Dukes' stage D patients was subjected to nested RT-PCR employing GC-C specific primers (Figure 1). Mononuclear cells were employed because preliminary studies confirmed that human colorectal cancer cells, like other epithelial tumor cells, co-segregate with mononuclear cells rather than with red cells, granulocytes, or platelets (data not shown). GC-C mRNA was detected in mononuclear cells from all 24 patients. However, identical results were obtained with mononuclear cells from all 20 volunteers, yielding a false positive rate of 100%.

Example 4: Identification of the Source of GC-C mRNA in Blood from Healthy Volunteers.

Blood from volunteers was separated into plasma, granulocytes, erythrocytes, platelets, and mononuclear cells, and 1 μg of total RNA from each of these components was analyzed by nested RT-PCR employing GC-C-specific primers (Figure 2A). GC-C mRNA was detected specifically in mononuclear cells, but not in other components of blood. To determine which cell population in the mononuclear cell fraction expressed GC-C transcripts, 1 μg of total RNA was extracted from purified monocytes, T cells, NK cells, B cells, and CD34+ cells (Bio-Whittaker; Charlotte, N.C.) and analyzed by nested RT-PCR employing GC-C-specific primers (Figure 2B). GC-C mRNA was detected specifically in CD34+ cells, but not in other purified cells.

Example 5: Analysis of Mononuclear Cells from a Breast Cancer Patient Undergoing Treatment with G-CSF.

Mononuclear cells were isolated from blood obtained on days 3 and 6 from a patient with breast cancer treated with 600 μg of G-CSF (Neupogen™ Amgen; Thousand Oaks, CA) on days 1 to 3. G-CSF stimulates the production of CD34+ progenitor cells and their mobilization from sites of hematopoiesis to the peripheral circulation, and is employed in stem cell harvests in preparation for bone marrow ablation and autologous transplantation. The leukocyte count for this patient was 107,000/μl on day 3 (final day of treatment with G-CSF) and 17,000/μl on day six (3 days following the last dose of G-

CSF). Total RNA (1 μg) extracted from mononuclear cells was subjected to RT-PCR, without nesting, employing GC-C specific primers (Figure 3). GC-C mRNA was detected in mononuclear cells obtained on day 3, when the concentration of CD34+ stem cells was maximal. In contrast, GC-C mRNA was not detected in mononuclear cells obtained on day 6, when the concentration of CD34+ stem cells had returned to baseline.

Example 6: Detection of GC-C mRNA Following Depletion of CD34+ Cells from Mononuclear Cells.

To further examine whether CD34+ progenitor cells were the source of GC-C mRNA in blood from healthy volunteers, mononuclear cells from these subjects were depleted of CD34+ cells and GC-C expression examined by nested RT-PCR. Total RNA (1 μg) extracted from mononuclear cells, CD34+-depleted mononuclear cells, and purified CD34+ cells was subjected to nested RT-PCR employing GC-C specific primers (Figure 4). GC-C mRNA was detected employing RNA from mononuclear cells and purified CD34+ cells, but not RNA from CD34+-depleted mononuclear cells.

Example 7: Expression of Epithelial Cell Markers in CD34+ Cells.

The utility of epithelial cell markers to detect rare circulating tumor cells in peripheral blood has been limited by high false positive rates reflecting illegitimate transcription in unknown blood components (12-20). The present studies demonstrate that CD34+ progenitor cells are the source of GC-C mRNA in the blood of healthy volunteers. These data suggest that CD34+ cells also may be the source of other illegitimately transcribed epithelial cell markers in blood. Thus, the expression of epithelial cell-specific transcripts, including CEA, PSA, PSM, CK-19, CK-20, MUC-1, and GA733.2 was examined in purified CD34+ cells (Fig 5). Total RNA (1 μg) extracted from purified CD34+ cells was subjected to nested RT-PCR employing primers specific for those epithelial cell markers (see Table 1). All epithelial cell-specific transcripts were expressed in RNA from CD34+ progenitor cells. As with GC-C, the expression of other epithelial cell markers was virtually eliminated by depleting mononuclear cells of CD34+ cells (Figure 6).

Example 8: Threshold for Detecting Illegitimate Transcripts of GC-C and CEA in Blood.

Total RNA from mononuclear cells of volunteers was serially diluted to define the threshold quantity of total RNA required to detect illegitimate transcripts of GC-C and CEA employing nested RT-PCR (Figure 7). GC-C and CEA transcripts were not detected employing <1 μg of RNA in any sample obtained from healthy volunteers (n=20). In contrast, GC-C mRNA was detected employing 0.1 μg of RNA and CEA amplicons were detected employing 0.5 μg of RNA from samples obtained from all Dukes' stage D patients (n=24) (Figure 8A). These data establish a threshold of total mononuclear cell RNA in RT-PCR reactions of ≥1 μg for detecting GCC and CEA transcripts arising from illegitimate transcription in CD34+ cells. Similarly, these data demonstrate that GC-C and CEA amplicons obtained in RT-PCR reactions employing ≤0.8 μg of total mononuclear cell RNA reflect the presence of circulating metastatic colorectal cancer cells.

Example 9: Sensitivity of Detecting Circulating Metastatic Colorectal Cancer Cells Employing GC-C Compared to CEA.

Total RNA from mononuclear cells obtained from Dukes' stage D patients was serially diluted and analyzed by RT-PCR employing GC-C- and CEA-specific primers (Figure 8B). As demonstrated above, GC-C and CEA transcripts detected employing <1 μg of RNA reflect circulating tumor cells in blood. CEA amplicons were detected in 7/24 (~30%) and 5/24 (~21%) Dukes' stage D patients employing 0.8 μg or 0.5 μg of RNA, respectively. In contrast, all (n=24) stage D patients yielded GC-C transcripts employing ≥0.1 μg of RNA.

Example 10: Sensitivity of Nested RT-PCR Employing GC-C-Specific Primers for Detecting Circulating Tumor Cells.

T84 or Caco2 human colon carcinoma cells (≈200 and 20 GC-C transcripts per cell, respectively) were serially diluted employing excess mononuclear cells, as indicated (Figure 9). Total RNA extracted from these samples (0.5 μg) was employed for nested RT-PCR employing GC-C-specific primers. A single T84 cell was detected in 10^7 mononuclear cells while one Caco2 cell was detected in 10^6 mononuclear cells. This level

of sensitivity for detecting human colorectal cancer cells by RT-PCR employing GC-C-specific primers was highly reproducible and yielded identical results when ten sequential analyses were performed.

Example 11: Disease Markers Associated with "Illegitimate Transcription."

The scientific literature describes a wide variety of diagnostic tests for disease markers whose diagnostic potential is compromised by background levels of "illegitimate transcription" in normal volunteers (Table 3). Most of these tests have either been abandoned as clinically useless or are considered less useful because of an inability to distinguish the background level from a true signal due to disease in patients. The tests for the markers described in Table 3 can be improved by removal, destruction or modification of CD34+ cells and/or other cells found to be the source of the background transcription.

Table 3. Disease markers with documented levels of "illegitimate transcription."

Marker	False Positives	Source	Reference
PSA (Prostate Specific Antigen)	4 %	Blood	1
	19 %	Blood	2
	35 %	Bone marrow	3
PSM (Prostate-specific membrane antigen)	96 %	Blood	2
	44 %	Bone Marrow	10
CK 18 (Cytokeratin 18)	71 %	Bone Marrow	10
	100 %	Blood	13
CK 19 (Cytokeratin 19)	23 %	Blood	4
	4-85 %	Blood	5
	71 %	Blood	6
	51 %	Blood	7
	60 %	Bone Marrow	7
	100 %	Lymph nodes	8
	64 %	Blood	9
	67 %	Bone marrow	9
	47 %	Mononucleated blood cells	9
	85 %	Peripheral blood stem cells	9
CK 20 (Cytokeratin 20)	50 %	Blood	14
	72 %	Blood	15
	6 %	Bone Marrow	16
CEA (Carcinoembryonic antigen)	26 %	Bone Marrow	10
	31 %	Lymph nodes	11
	33 %	Blood	12
Erb-B2	71 %	Bone Marrow	10
Erb-B3	86 %	Bone Marrow	10
MUC-1 (Epithelial mucin-1)	46 %	Blood	5
	44 %	Lymph nodes	17

	100 %	Lymph nodes	5
MUC-18 (Epithelial mucin-18)	5 %	Blood	17
GA 733.2 (gastrointestinal tumor-associated antigen 733.2)	54 %	Blood	5
	100 %	Lymph nodes	5
DPL I (desmoplakin I)	100 %	Bone Marrow	5
EGP-40 (Epithelial glycoprotein 40)	100 %	Bone marrow	5
Tyrosinase	50 %	Lymph nodes	18
TGB (Thyroglobulin)	10 %	Blood	19
Tyrosine Hydroxylase	14 %	Blood	20
	57 %	Bone Marrow	20
NPGP 9.5 (Neuron-specific glycoprotein)	57 %	Blood	21

1. Lehrer *et al.*, 1996, Br. J. Cancer, 74:871-873.
2. Gala *et al.*, 1998, Clin. Chem., 44:472-481.
3. Albers *et al.*, 2000, Anticancer Res., 20:2107-2112.
4. Battaglia *et al.*, 1998, Bone Marrow Trans., 22:693-698.
5. Bostick *et al.*, 1998, J. Clin. Oncol., 16:2632-2640.
6. Traystman *et al.*, 1997, J. Hematother., 6:551-561.
7. Slade *et al.*, 1999, J. Clin. Oncol., 17:870-879.
8. Schoenfeld *et al.*, 1994, Cancer Res., 54:2986-2990.
9. Lopez-Guerrero *et al.*, 1997, Clin. Chem. Acta, 263:105-116.
10. Zippelius *et al.*, 1997, J. Clin. Oncol., 15:2701-2708.
11. Mori *et al.*, 1995, Cancer Res., 55:3417-3420.
12. Ko *et al.*, 1998, Clin. Cancer Res., 4:2141-2146.
13. Brown *et al.*, 1995, Surgery, 117:95-101.
14. Hardingham *et al.*, 2000, Int. J. Cancer, 89:8-13.
15. Champelovier *et al.*, 1999, Anticancer Res., 19:2073-2078.
16. Soeth *et al.*, 1996, Int. J. Cancer, 69:278-282.
17. Hoon *et al.*, 1995, Cancer, 76:533-534.
18. Calogero *et al.*, 2000, Br. J. Cancer, 83:184-187.
19. Ringel *et al.*, 1998, J. Clin. Endocrinol. Metab., 83:4435-4442.
20. Miyajima *et al.*, 1996, Cancer, 77:1214-1219.
21. Mattano *et al.*, 1992, Cancer Res., 52:4701-4705.